Linear Regression: Prediction Plots, Planning

STAT 245

Sept 16-18, 2024

Prediction Plots

Vary only *certain* predictor(s)

- **We can't just show "data plus line"** with multiple predictors
- New dataset with desired predictor values

New Prediction Data

```
fake_data <-
  expand.grid(fWHR = seq(from = _____,
                         to = _____,
                         by = _____),
              normDS = __________,
              Sex = _________,
              Group = ___________)
```

Hypothetical Data

Quantitative Predictor

One predictor varies; the others are held constant at median or most common or common-sense values (don't include impossible combinations!)

Hypothetical Data

Quantitative Predictor

```
library(mosaic) # for mean()
fake_data <- expand_grid(
  fWHR = seq(from = 1.05, by = 0.01, to =
1.9),
  normDS = mean(~normDS,
                data = bonobos,
                na.rm = TRUE),
  Sex = 'Female',
  Group = 'Planckendael')
```

Make Predictions

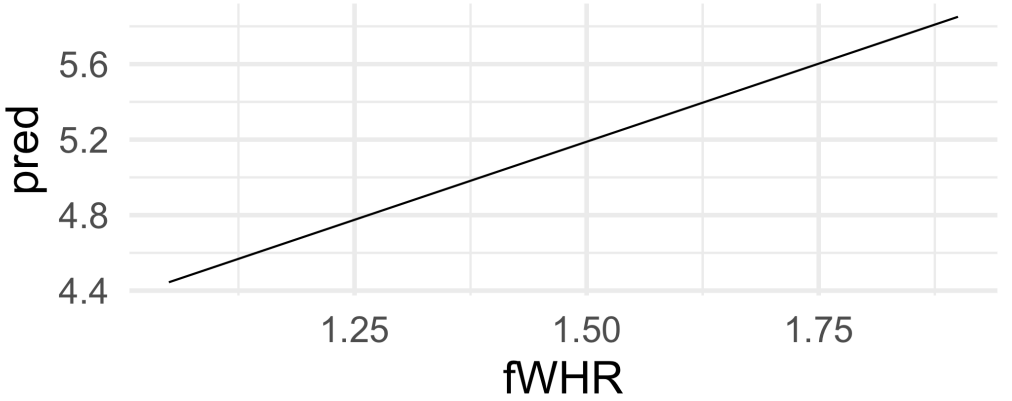
```
fake_data <- fake_data |>
  mutate(pred = predict(m3_2q2c,
                        newdata = fake_data))
glimpse(fake_data)
```

```
## Rows: 86
## Columns: 5
## $ fWHR   <dbl> 1.05, 1.06, 1.07, 1.08, 1.09, 1.10, 1.11, 1.12, 1.13, 1.14,
1.1…
## $ normDS <dbl> 2.657017, 2.657017, 2.657017, 2.657017, 2.657017, 2.657017,
2.6…
## $ Sex    <chr> "Female", "Female", "Female", "Female", "Female", "Female",
"Fe…
## $ Group  <chr> "Planckendael", "Planckendael", "Planckendael",
"Planckendael",…
## $ pred   <dbl> 4.444305, 4.460851, 4.477397, 4.493943, 4.510489, 4.527036,
4.5…
```

Prediction Plots

Create the Graph

```
gf_line(pred ~ fWHR, data = fake_data)
```



BUT What is still missing?

Uncertainty!

On predictions: a confidence interval gives a range of plausible values for average response, taking into account uncertainty in intercept and slope estimates.

Relying on the Central Limit Theorem, a simple CI is:

$$\text{estimate} \pm 1.96 * \text{standard error}$$

SE for predictions

Should account for uncertainty in *all* the βs

```
preds <- predict(m3_2q2c,
                 newdata = fake_data,
                 se.fit = TRUE)
glimpse(preds)
```

```
## List of 4
##  $ fit           : Named num [1:86] 4.44 4.46 4.48 4.49 4.51 ...
##   ..- attr(*, "names")= chr [1:86] "1" "2" "3" "4" ...
##  $ se.fit        : Named num [1:86] 0.304 0.3 0.296 0.293 0.289 ...
##   ..- attr(*, "names")= chr [1:86] "1" "2" "3" "4" ...
##  $ df            : int 107
##  $ residual.scale: num 0.923
```

Put Preds + SEs *in dataset*

```
fake_data <- fake_data |>
  mutate(pred = preds$fit,
         pred.se = preds$se.fit)
```

Result?

```
glimpse(fake_data)
```

```
## Rows: 86
## Columns: 6
## $ fWHR    <dbl> 1.05, 1.06, 1.07, 1.08,
1.09, 1.10, 1.11, 1.12, 1.13, 1.14, 1.…
## $ normDS  <dbl> 2.657017, 2.657017,
2.657017, 2.657017, 2.657017, 2.657017, 2.…
## $ Sex     <chr> "Female", "Female",
"Female", "Female", "Female", "Female", "F…
## $ Group   <chr> "Planckendael",
"Planckendael", "Planckendael",
```

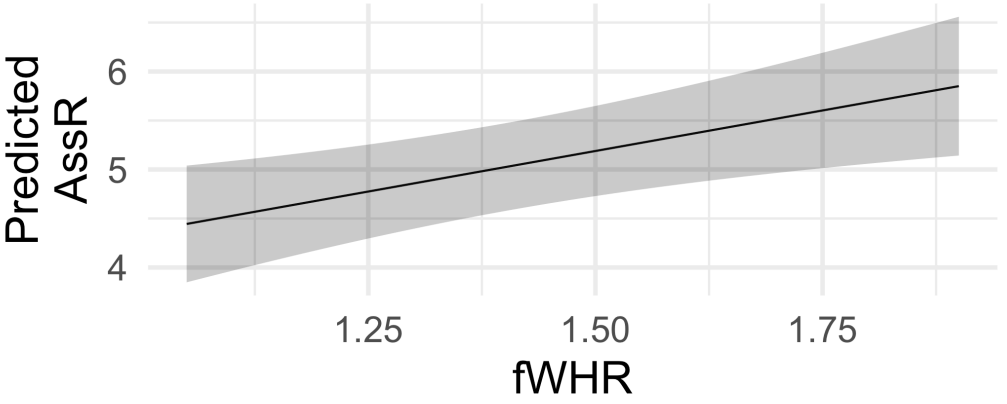
From SE to CI

```
fake_data <- fake_data |>
  mutate(CI_lower = pred - 1.96*pred.se,
         CI_upper = pred + 1.96*pred.se)
glimpse(fake_data)
```

```
## Rows: 86
## Columns: 8
## $ fWHR     <dbl> 1.05, 1.06, 1.07, 1.08, 1.09, 1.10, 1.11, 1.12, 1.13,
1.14, 1…
## $ normDS   <dbl> 2.657017, 2.657017, 2.657017, 2.657017, 2.657017,
2.657017, 2…
## $ Sex      <chr> "Female", "Female", "Female", "Female", "Female",
"Female", "…
## $ Group    <chr> "Planckendael", "Planckendael", "Planckendael",
"Planckendael…
## $ pred     <dbl> 4.444305, 4.460851, 4.477397, 4.493943, 4.510489,
4.527036, 4…
## $ pred.se  <dbl> 0.3037098, 0.3000145, 0.2963808, 0.2928109, 0.2893073,
0.2858…
## $ CI_lower <dbl> 3.849034, 3.872823, 3.896491, 3.920034, 3.943447,
```

Plot Pred. w/CI

```
gf_line(pred ~ fWHR,
        data = fake_data) |>
  gf_labs(y='Predicted\nAssR') |>
  gf_ribbon(CI_lower + CI_upper ~ fWHR)
```



Categorical Predictors?

Replace `lines` with `points` and `ribbon` with `errorbar`

- new fake data
- slightly different plotting code

Hypothetical Data

Categorical Predictor

```
fake_data <-
  expand.grid(fWHR = 1.4,
              normDS = 2.4,
              Sex = c('Female', 'Male'),
              Group = 'Planckendael')
```

Make Predictions

Categorical Predictor

```
preds <- predict(m3_2q2c,
                 newdata = fake_data,
                 se.fit = TRUE)
glimpse(preds)
```

```
## List of 4
##  $ fit           : Named num [1:2] 5.01 3.77
##   ..- attr(*, "names")= chr [1:2] "1" "2"
##  $ se.fit        : Named num [1:2] 0.226 0.219
##   ..- attr(*, "names")= chr [1:2] "1" "2"
##  $ df            : int 107
##  $ residual.scale: num 0.923
```

Convert to CI

Categorical Predictor

```
fake_data <- fake_data |>
  mutate(pred = preds$fit,
         pred.se = preds$se.fit,
         CI_lower = pred - 1.96*pred.se,
         CI_upper = pred + 1.96*pred.se)
glimpse(fake_data)
```

```
## Rows: 2
## Columns: 8
## $ fWHR     <dbl> 1.4, 1.4
## $ normDS   <dbl> 2.4, 2.4
## $ Sex      <fct> Female, Male
## $ Group    <fct> Planckendael, Planckendael
## $ pred     <dbl> 5.005257, 3.771274
## $ pred.se  <dbl> 0.2260898, 0.2191601
## $ CI_lower <dbl> 4.562121, 3.341720
## $ CI_upper <dbl> 5.448393, 4.200828
```

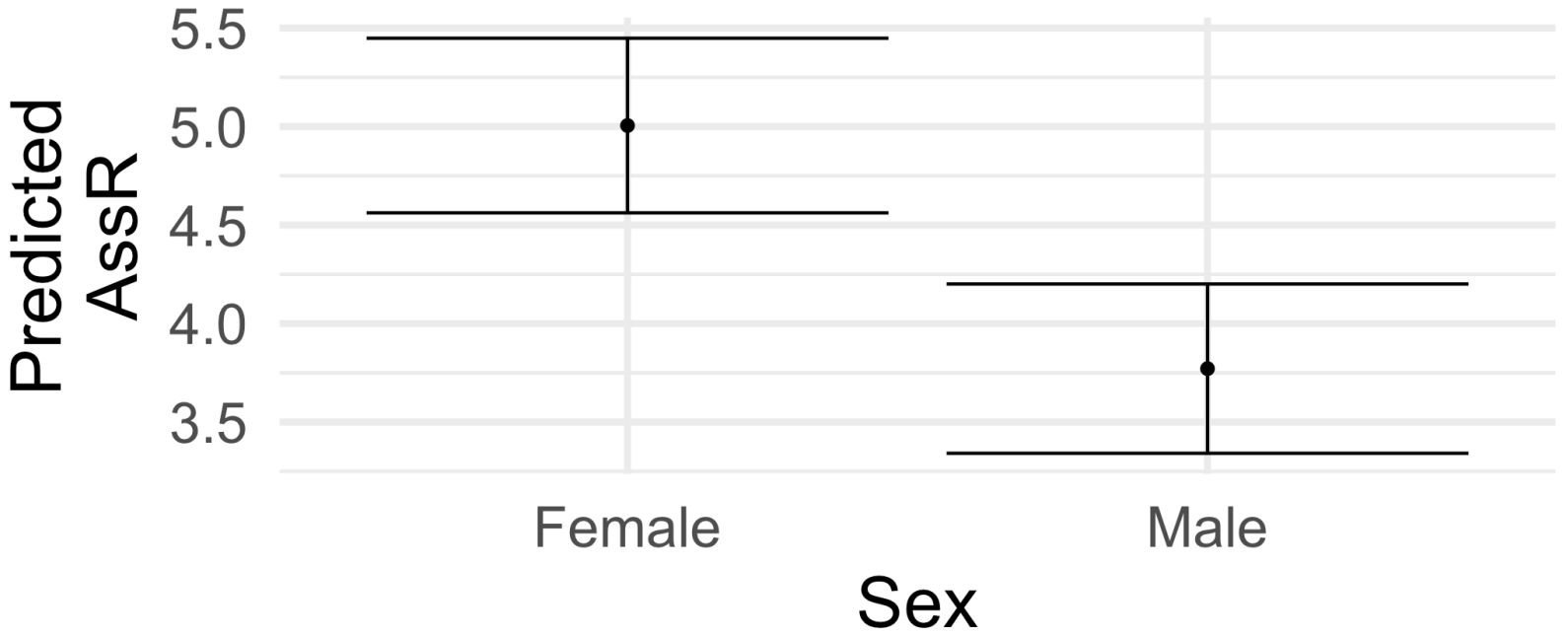
Prediction Plot (code)

Categorical Predictor

```
gf_point(pred ~ Sex,
         data = fake_data) |>
  gf_labs(y='Predicted\n AssR') |>
  gf_errorbar(CI_lower + CI_upper ~ Sex)
```

The Prediction Plot

Categorical Predictor



R so far: for data

- `▷` (pipe) for "and then..."
- `mutate()` to add variable to dataset
- `select()` to keep certain variables
- `na.omit()` to remove rows w/missing data (!!)
- `glimpse()` to peek at dataset
- *`pander::pander()` to print table*

R so far: for models

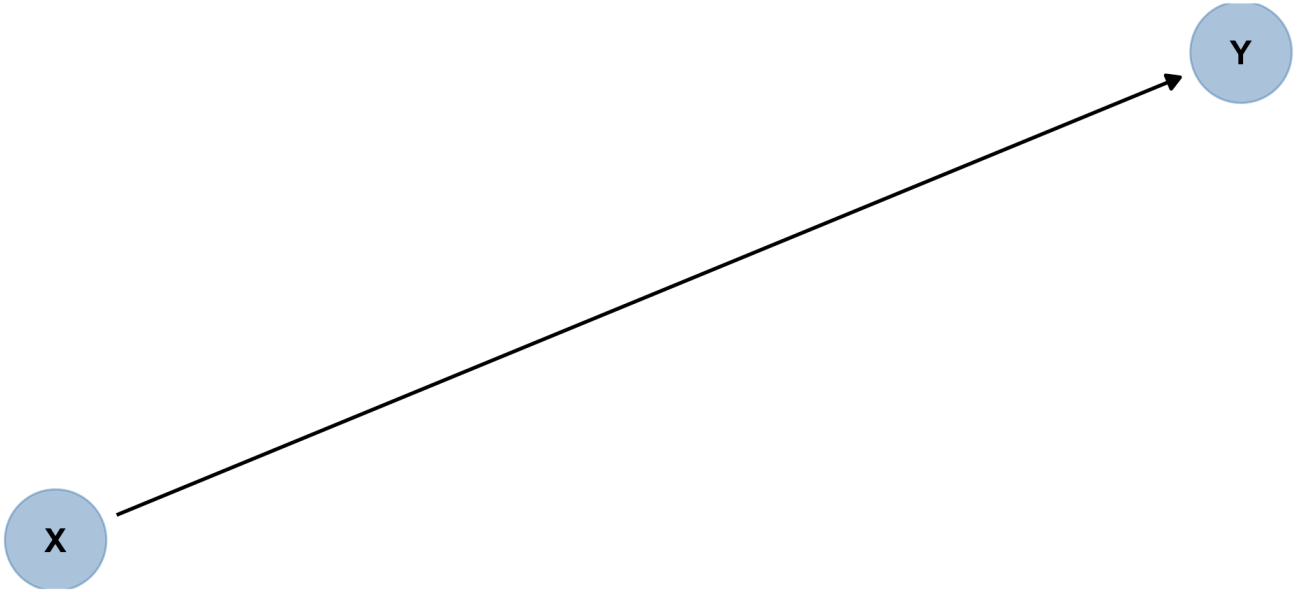
- `lm(y ~ x1 + x2, data = ___)` to fit linear model
- `resid(model)` to
- `predict(model, ...)` for prediction
  - `se.fit = TRUE` (or `FALSE`)
  - `newdata = ...`

R so far: graphics

- `gf_ribbon()` to add error band
- `gf_errorbar()` to add error bars

Causal Diagrams

There's more to planning than just p < n/15!



PREKNOP Example

Response: Knowledge of Body (KoB) Score

- Parity
- Wish to conceive
- Before/After course
- "Pre" KoB score
- Age
- Education
- Race/Ethnicity
- Income
- Health Insurance

Confounder

Precision Covariate

Mediator

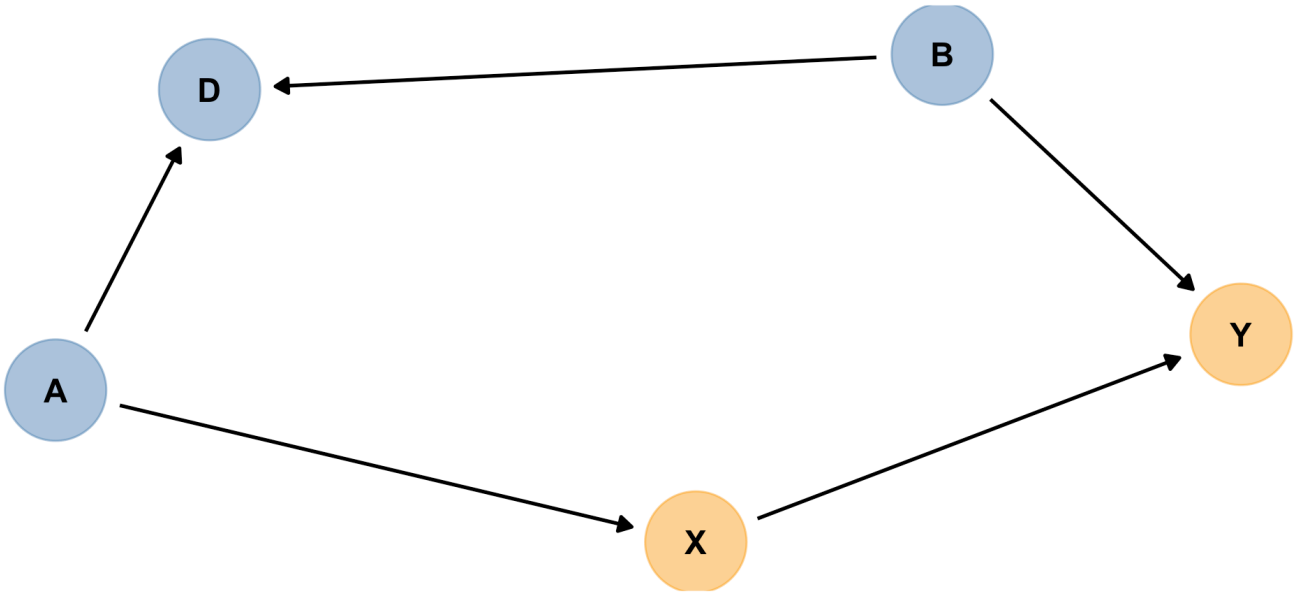
Moderator or Modifier

Also known as: *Interaction*

Collider

Can get complex

M-Bias



Look back: which variables are "in" a PREKNOP model?

Resource: Guide to Causal Inference https://doi.org/10.1098/rspb.2020.2815

Your Summary

Linear modeling step-by-step: